



## DE BASIS voor vindbaarheid

DE BASIS gebruikt het begrip 'vindbaarheid' voor de situatie waarin digitale objecten en bijbehorende metadata zodanig worden aangeboden, dat zij voor mens en machine bereikbaar, uniek identificeerbaar en herbruikbaar zijn. Met DE BASIS voor vindbaarheid zorg je ervoor dat je digitale collecties (waarin vaak fors is geïnvesteerd) toegankelijk zijn voor een groter en breder publiek.



Wanneer je DE BASIS voor vindbaarheid toepast zorg je ervoor dat digitale objecten gevonden kunnen worden door zoekmachines, er probleemloos naar digitale objecten verwezen kan worden en dat objecten aangeboden kunnen worden op andere platformen dan je eigen website (denk aan initiatieven als [Europeana](#) en andere regionale, thematische of nationale [portals](#)). Tenslotte vormt DE BASIS voor vindbaarheid een eerste stap om de digitale collecties aan te bieden als [linked \(open\) data](#).

### VIER SLEUTELBEGRIPPEN

We delen 'vindbaarheid' in DE BASIS op in vier sleutelbegrippen:

1. **Identificatie** van data: de data zijn uniek en liefst duurzaam identificeerbaar.
2. **Bereikbaarheid** van data: de data zijn bereikbaar via internet.
3. **Zoekmachinedoorzoekbaarheid** van data: de data worden zo aangeboden dat zij doorzoekbaar zijn voor zoekmachines.
4. **Hergebruik** van data: de data kunnen (eenvoudig) worden geharvest, hergebruikt door derden en - via linked data - gekoppeld aan andere datasets.

*Het woord 'data' staat hierbij zowel voor de digitale objecten als de metadata.*

Wanneer je deze vier stappen van begin tot einde doorloopt, ontstaat er een optimale vindbaarheid van de digitale collecties.

### 1. IDENTIFICATIE VAN DATA

De eerste stap naar vindbaarheid is het uniek identificeerbaar maken van het digitale (en eventueel bijbehorende analoge) object en de bijbehorende metadata. Dit proces is opgedeeld in drie 'fases':

1. In de eerste fase voorzie je elk digitaal object van een unieke identifier. Dit houdt in dat je elk digitaal object – binnen de instelling – een uniek nummer of code geeft.
2. In de tweede fase maak je een URI ([Uniform Resource Identifier](#)) aan voor elk digitaal object. Dit is een unieke verwijzing naar een digitaal object. Op internet vind je een URI vaak terug in de vorm van een URL (Uniform Resource Locator, of wel een

locatieverwijzing op de server), een URN (een naamsverwijzing, niet verbonden aan een fysieke plek op een server) of een combinatie van beiden. DE BASIS schrijft het gebruik van URI's voor. Met een URI kun je een begin maken met het aanbieden van je data als [linked \(open\) data](#).

3. In de derde fase zorg je er voor dat de URI duurzaam, ofwel persistent, is. De URI is dan losgekoppeld van de daadwerkelijk locatie van de digitale bron. Achter de schermen zorgt een tabel ervoor dat de daadwerkelijke, aan verandering onderhevige, locatie wordt gekoppeld aan de onveranderlijke URI. Dit wordt *resolven* of *redirecten* genoemd. Om persistente identificatie te bereiken zijn er veel methodes en standaarden beschikbaar:

- [Handle](#)
- [PURL](#)
- [OpenURL](#)
- [ARK](#)

DE BASIS schrijft geen van deze standaarden dwingend voor, maar beveelt in ieder geval het gebruik van een van de bovengenoemde standaarden aan.

## 2. BEREIKBAARHEID VAN DATA

Om de data op internet bereikbaar te maken, moeten de data getransporteerd kunnen worden met behulp van het [HTTP](#)- (of het beveiligde [HTTPS](#)-) protocol. Wanneer je het downloaden van hele datasets tegelijk mogelijk wil maken, ligt het [FTP](#)-protocol meer voor de hand.

## 3. ZOEKMACHINEDOORZOEKBAARHEID VAN DATA

Om de digitale collecties (dat wil zeggen: de website én de 'achterliggende' databases) bereikbaar te maken voor zoekmachines, zijn een aantal extra stappen nodig. Een minimum vereiste is dat voor ieder digitaal object een 'HTML landingspagina' (ook wel *landing page*) beschikbaar moet zijn. Wanneer elk digitaal object een eigen landing page heeft, kan die door zoekmachines worden gevonden. Dit kan een statische pagina zijn, maar een dynamisch gegenereerde pagina is daarvoor ook geschikt.

Om de hele digitale collectie vervolgens aan te bieden aan zoekmachines zijn er twee mogelijkheden die door DE BASIS worden aanbevolen (en naast elkaar gebruikt kunnen worden):

1. Aanbieden via Sitemaps: een volgens het protocol [sitemap.org](#) opgestelde lijst van URL's en metadata (=sitemap) biedt de zoekmachines de mogelijkheid de hele website, inclusief de achterliggende databases, te indexeren.
2. Aanbieden via hypertext links: als alle URL's (die naar individuele objecten uit de digitale collectie verwijzen) bereikbaar zijn via hyperlinks, kan een zoekmachine deze URL's 'spideren'. Dit kan bijvoorbeeld door de collectie onder te verdelen in rubrieken, subrubrieken, etc., en deze aanklikbaar te maken, om zo te kunnen bladeren door de collectie.

Zie voor verdere optimalisatie van webpagina's voor zoekmachines ook de [DE BASIS voor presentatie](#).

## 4. HET GEBRUIK VAN DATA

Om hergebruik van data mogelijk te maken zijn er in DE BASIS voor vindbaarheid drie eisen opgenomen:

1. Om je data uitwisselbaar te maken, zou je deze in ieder geval in een gestructureerd en open formaat aan moeten bieden. Het gebruik van (betekenisvolle) XML (bij voorkeur valideerbaar, door middel van een XML Schema definitie bestand: een [XSD file](#)) is een voor de hand liggende manier om de data te structureren. DE BASIS schrijft dit als minimum standaard voor.
2. Om je data vervolgens te kunnen ophalen uit het informatiesysteem zonder arbeidsintensieve exportacties, heb je [OAI-PMH](#) (Open Archives Initiative Protocol for Metadata Harvesting) nodig. Deze standaard maakt het mogelijk dat data uit verschillende bronnen zonder tussenkomst van mensen 'geogst' kunnen worden. Er is dan sprake van de 'aggregatie' van data door een ander informatiesysteem.
3. De standaard [Dublin Core Metadata Element Set](#) maakt erfgoedbeschrijvingen uitwisselbaar door een (in beginsel) beperkte set van 15 sectoroverstijgende standaardvelden (ook wel: Simple Dublin Core) te hanteren. OAI-PMH (in combinatie met Dublin Core) is een belangrijke voorwaarde om je data beschikbaar te stellen aan portals of aggregators als Europeana en wordt daarom door DE BASIS als minimum standaard voorgeschreven.

Alternatieve en gespecialiseerde uitwisselbare metadatastandaarden zoals [EAD](#) (voor archiefcollecties) en [LIDO](#) (voor museumcollecties) worden behandeld in het [kennisdossier Vindbaarheid](#) en maken niet expliciet deel uit van DE BASIS.

Naast deze eisen doet DE BASIS twee aanbevelingen voor hergebruik van data:

1. Een andere mogelijkheid om data te delen, is door gebruik te maken van [RESTful API's](#) (een webservice die werkt volgens het [REST-principe](#)). Met behulp van de API kan een volledige of gedeeltelijke dataset worden hergebruikt in een andere context. Het gebruik van een RESTful API wordt door DE BASIS aanbevolen.
2. Wanneer je het hergebruik van je data nog verder wilt vergroten kun je je data aanbieden als [RDF](#) (Resource Description Framework). Objecten die zijn beschreven in RDF kunnen via zogenaamde [triples](#) betekenisvolle relaties aangaan met andere objecten. RDF is daarmee de basis voor het aanbieden van informatie als [linked \(open\) data](#). RDF is voorbeeld van een *better practice* (in tegenstelling tot een *basic practice*) en maakt daarom niet expliciet deel uit van DE BASIS.

### Licenties voor hergebruik van data

Het hergebruik van data heeft ook een juridische component. In hoeverre kunnen de data worden hergebruikt buiten de context van de eigen instelling? Op deze en andere vragen over de voorwaarden van het gebruik van data wordt ingegaan in het [kennisdossier Rechtenbeheer](#). Bovendien wordt er 2013 een aparte BASIS voor auteursrecht ontwikkeld.

## SAMENVATTEND

De volgende standaarden zijn een vereiste om te voldoen aan DE BASIS voor vindbaarheid:

- **Identificatie van data:**  
Het gebruik van een persistente URI. Gebruik daarvoor een van de beschikbare standaarden of methodes, zoals Handle, PURL of OpenURL.
- **Bereikbaarheid van data op het internet:**  
Biedt de data aan via HTTP(s) of FTP.
- **Zoekmachinedoorzoekbaarheid van data:**  
Het aanmaken van een HTML-landingspagina voor elk digitaal object. Als aanbeveling hiervoor geldt het aanbieden van een sitemap opgesteld volgens protocol van [sitemap.org](http://sitemap.org) en/of het aanbieden van alle via URL's bereikbare digitale objecten via hypertext links.
- **Hergebruik van data:**  
Het aanbieden van data in een gestructureerd en open formaat (lees: betekenisvolle XML-valideerbaar door een XML Schema definitie bestand (XSD)), OAI-PMH en Simple Dublin Core. Het gebruik van RESTful API 's voor het aanbieden van data via een webservice geldt als aanbeveling.

## MEER WETEN?

- Lees het [kennisdossier Vindbaarheid](#) en de onderliggende kennisdossiers.
- Lees de [samenvatting](#) (pdf) van de rapportage over de totstandkoming van de voormalige DE BASIS voor vindbaarheid (2007).

## VERANTWOORDING

Dit is een herziene, tweede versie van DE BASIS voor vindbaarheid (maart 2013). De eerste versie stamt uit 2007. Deze herziene tekst is opgesteld door [experts](#) die werkzaam zijn op het gebied van metadata, linked open data, harvesting, aggregatie van digitaal erfgoedmateriaal en in overleg met de [adviesraad van de kennisbank van DEN](#).

## DENK MEE!

Deze tekst staat altijd open voor commentaar. Wil je reageren of meedenken? Maak dan gebruik van het onderstaande reactieformulier of lever je commentaar rechtsreeks in de tekst via het bijgevoegde MSWord- of PDFbestand. Stuur deze s.v.p. naar [den@den.nl](mailto:den@den.nl) (onder vermelding van *DE BASIS voor vindbaarheid*).